



Bayesian estimation of the tail index of a heavy tailed distribution under random censoring

Abdelkader Amaraoui, Kamal Boukhetala, Jean-François Dupuy

► To cite this version:

Abdelkader Amaraoui, Kamal Boukhetala, Jean-François Dupuy. Bayesian estimation of the tail index of a heavy tailed distribution under random censoring. *Computational Statistics and Data Analysis*, 2016, 104, pp.148-168. 10.1016/j.csda.2016.06.009 . hal-01071496

HAL Id: hal-01071496

<https://hal.science/hal-01071496>

Submitted on 14 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bayesian estimation of the tail index of a heavy tailed distribution under random censoring

Abdelkader Ameraoui¹, Kamal Boukhetala

Faculty of mathematics PoBox 32, Al alia Bab ezzouar, USTHB - Algiers, Algeria
and

Jean-François Dupuy

IRMAR-INSA de Rennes, 20 Avenue des Buttes de Coësmes, 35708 Rennes cedex 7,
France

Abstract

Bayesian estimation of the tail index of a heavy-tailed distribution is addressed when data are randomly right-censored. Maximum a posteriori and mean posterior estimators are constructed for various prior distributions of the tail index and their consistency and asymptotic normality are established. Finite-sample properties of the proposed estimators are investigated via simulations. Tail index estimation requires selecting an appropriate threshold for constructing relative excesses. A Monte Carlo procedure is proposed for tackling this issue. Finally, the proposed estimators are illustrated on a medical dataset.

Keywords: Extreme value theory, asymptotic properties, maximum a posteriori estimator, mean posterior estimator, simulations.

1. Introduction

Tail index estimation is one of the most important issues in extreme value theory. The tail index measures the thickness of the tail of a probability distribution function and thus plays a crucial role for evaluating the risk of occurrence of extremes events. In particular, estimation of the tail index constitutes usually a first step in an extreme value analysis. A vast literature has been dedicated to this topic. Recent overviews can be found in the monographs [2] and [17].

Let F be the cumulative distribution function (cdf) of some non-negative random variable Y . We assume that F is heavy-tailed, that is, there exists a constant $\alpha > 0$ such that

$$1 - F(x) = x^{-\alpha}\ell(x), \quad (1)$$

where ℓ is a slowly varying function at infinity:

$$\lim_{x \rightarrow \infty} \frac{\ell(tx)}{\ell(x)} = 1 \quad \text{for all } t > 0.$$

If (1) holds, we have:

$$\lim_{x \rightarrow \infty} \frac{1 - F(tx)}{1 - F(x)} = t^{-\alpha} \quad \text{for all } t > 0$$

and we say that F is regularly varying at infinity with tail index α , which we denote by $F \in \mathcal{R}_{-\alpha}$. The positive number $\gamma := \alpha^{-1}$ is called the extreme value index (EVI) of F . The conditions above amount to assuming that the distribution function F is in the max-domain of attraction of a Fréchet distribution. Such distribution functions are useful in practice for investigating phenomena where exceptional

¹Corresponding author : aameraoui@usthb.dz

values have a significant occurrence frequency. Examples include the number of claims in insurance [12], transmission times in telecommunications [25], log-returns of price speculation [12].

Several estimators have been proposed for the tail index α , or equivalently, for the EVI γ , such as Pickands estimator [23], Dekkers *et al.* (or moment) estimator [8] and Hill's estimator [18], which is the most popular estimator of γ in model (1). Let X_1, X_2, \dots, X_n be independent and identically distributed (iid thereafter) random variables with common cdf F . Let $k \in \{2, \dots, n\}$ and $X_{n,1} \leq X_{n,2} \leq \dots \leq X_{n,n}$ be the ordered statistics associated to the sample X_1, X_2, \dots, X_n . Hill's estimator is defined as

$$H(k) := \frac{1}{k} \sum_{i=1}^k \log(X_{n,n-i+1}) - \log(X_{n,n-k}). \quad (2)$$

Consistency of Hill's estimator was proved in [21] under the regular variation condition (1). Its asymptotic normality was further established under an additional condition known as the second-order regular variation condition (see [17], page 117).

In this paper, we address estimation of the tail index α when observations X_1, X_2, \dots, X_n are randomly right-censored. Censoring commonly occurs in the analysis of event time data. For example, X may represent the duration until the occurrence of some event of interest, such as death of a patient. If a patient is still alive or has dropped out of the study for some reason when the data are collected, the variable of interest X is not available. An appropriate way to model this situation is to introduce a random variable Y (called a censoring random variable) such that the observations consist of pairs (Z_i, δ_i) , $1 \leq i \leq n$ where $Z_i = \min(X_i, Y_i)$, $\delta_i = \mathbf{1}_{\{X_i \leq Y_i\}}$ and $\mathbf{1}$ is the indicator function. Estimation of the EVI with censored data was considered in [3], [4], [11], [15] and [16]. For example, Beirlant *et al.* (2007) [3] proposed to estimate γ by the following modified version of Hill's estimator:

$$H^C(k) := \frac{\sum_{i=1}^k \log(Z_{n,n-i+1}) - \log(Z_{n,n-k})}{\sum_{i=1}^k \delta_{[n-i+1]}}, \quad (3)$$

where $Z_{n,1} \leq Z_{n,2} \leq \dots \leq Z_{n,n}$ are the ordered statistics associated to the censored sample Z_1, Z_2, \dots, Z_n and $\delta_{[n-i+1]}$ is the concomitant value of δ associated with $Z_{n,n-i+1}$. More generally, Einmahl *et al.* (2008) [11] and Gomes and Neves (2011) [15] proposed to estimate γ by

$$\hat{\gamma}_Z^{(C)}(k) = \frac{\hat{\gamma}_Z(k)}{\hat{p}}, \quad (4)$$

where $\hat{\gamma}_Z(k)$ is any of the classical EVI estimators calculated on the censored observations Z_1, \dots, Z_n and $\hat{p} = \frac{1}{k} \sum_{i=1}^k \delta_{[n-i+1]}$ is the proportion of uncensored values in the k largest observations of Z . Obviously, the estimator (4) coincides with the adapted Hill's estimator (3) when $\hat{\gamma}_Z(k)$ is Hill's estimator (2). In this paper, we adopt a completely different approach and investigate Bayesian estimation of the tail index $\alpha := \gamma^{-1}$.

Bayesian estimation provides an alternative to frequentist methods. In the context of extreme values analysis without censoring, Bayesian estimators have been investigated in [5], [6], [10], [22], [27], [30]. See also [2] (chapter 11). Applications include operational risk modeling [13], hydrology [20] and market indices modeling [26]. But to the best of our knowledge, no Bayesian estimator of the tail index α has been proposed when censoring is present. The present work intends to fill this gap.

We construct several Bayesian estimators of α in model (1). Bayesian estimation requires specifying a prior distribution for the unknown parameter. We investigate various priors (Jeffreys, maximal data information and conjugate Gamma priors), leading to several maximum a posteriori and mean posterior estimators of α . We establish consistency of these estimators and asymptotic normality of their posterior distribution. Their finite-sample performance are assessed via simulations. Tail index estimation requires choosing an appropriate threshold for defining excesses. We propose a randomization method for tackling this issue. This procedure is evaluated by simulations. Finally, we illustrate our methodology on a set of AIDS survival data.

The paper is organized as follows. In Section 2, we construct our estimators and we establish their asymptotic properties. All proofs are deferred to an appendix. Section 3 reports the results of a comprehensive simulation study. The randomization procedure of the threshold is proposed and assessed by

simulations in Section 4. Application to AIDS data is carried out in Section 5. A discussion and some perspectives are given in Section 6.

2. Bayesian estimation of the tail index

We first introduce some notations. Let X_1, X_2, \dots, X_n be n iid copies of a non-negative random variable X with cdf F . The probability density function of X is denoted by f . We assume that F is heavy-tailed with tail index α , *i.e.* $F \in \mathcal{R}_{-\alpha}$. Let Y_1, Y_2, \dots, Y_n be an iid sample of the non-negative censoring random variable Y with cdf G and density function g . We assume that G is also heavy-tailed, with positive tail index β , *i.e.* $G \in \mathcal{R}_{-\beta}$. We assume that X and Y are independent and that we observe the n independent pairs

$$(Z_i, \delta_i), 1 \leq i \leq n,$$

where $Z_i = \min(X_i, Y_i)$ and $\delta_i = \mathbf{1}_{\{X_i \leq Y_i\}}$. We let $Z_{n,1} \leq Z_{n,2} \leq \dots \leq Z_{n,n}$ be the ordered statistics associated to the sample (Z_1, Z_2, \dots, Z_n) and $\delta_{[n-i+1]}$ be the concomitant value of δ associated with $Z_{n,n-i+1}$, that is, $\delta_j = \delta_{[n-i+1]}$ if $Z_j = Z_{n,n-i+1}$. Let H be the cdf of Z . Note that by independence of X and Y , H is also heavy-tailed and $H \in \mathcal{R}_{-(\alpha+\beta)}$.

Now, we define the random variables $E_{j,u} = \frac{Z_j}{u}$ given $Z_j > u$, which represent relative excesses over the threshold u . Let N_u denote the number of such excesses. Then we can obtain the partial likelihood of α based on the sample $(E, \Delta) = (e_{j,u}, \delta_{[n-N_u+j]})_{j=1, \dots, N_u}$ (see [1], [3]):

$$\begin{aligned} \mathcal{L}(E, \Delta | \alpha) &= \prod_{j=1}^{N_u} \left[\alpha e_{j,u}^{-(\alpha+1)} \right]^{\delta_{[n-N_u+j]}} \left[e_{j,u}^{-\alpha} \right]^{1-\delta_{[n-N_u+j]}} \\ &= \alpha^{\sum_{j=1}^{N_u} \delta_{[n-N_u+j]}} \left(\prod_{j=1}^{N_u} e_{j,u} \right)^{-\alpha} \left(\prod_{j=1}^{N_u} e_{j,u}^{-\delta_{[n-N_u+j]}} \right). \end{aligned} \quad (5)$$

In the Bayesian framework, unknown parameters are considered as random variables whose probability density function (the so-called prior density) represent the initial set of beliefs about the parameters. The prior distribution is updated by using the information contained in the data, yielding the so-called posterior density of the parameters. In our setting, we provide the unknown tail index α with a prior density $\pi(\alpha)$. Then, using Bayes' theorem, we obtain the posterior density $\pi(\alpha | E, \Delta)$ of α , which writes as:

$$\pi(\alpha | E, \Delta) = \frac{\mathcal{L}(E, \Delta | \alpha) \pi(\alpha)}{\int_{\Lambda} \mathcal{L}(E, \Delta | \alpha) \pi(\alpha) d\alpha},$$

where Λ is the support of the distribution of α and the denominator $\int_{\Lambda} \mathcal{L}(E, \Delta | \alpha) \pi(\alpha) d\alpha$ is a normalization constant independent of α . The posterior distribution of α is proportional to the product of the partial likelihood (5) and the prior, namely:

$$\pi(\alpha | E, \Delta) \propto \mathcal{L}(E, \Delta | \alpha) \pi(\alpha).$$

Choosing the prior density is a central issue in Bayesian estimation. When the information available for prior elicitation is minimal, one can use objective (or non-informative) priors, such as Jeffreys prior [19] and the maximal data information (MDI) prior introduced by Zellner (1971) (see [31] and [32]). On the other hand, when prior choice can be based on experts opinion, so-called subjective prior distributions can be used, such as conjugate priors. Conjugate priors are such that the posterior distribution of the parameter is in the same family as the prior. In Subsections 2.1 and 2.2, we investigate MDI and Jeffreys priors for the estimation of the tail index α under random censoring. In Subsection 2.3, we investigate a conjugate Gamma prior.

2.1. Maximal data information (MDI) prior

The MDI prior was defined so as to maximize the average information in the data density relative to that in the prior (see [31]). If $\mathcal{L}(E, \Delta|\alpha)$ is the likelihood of a single observation (E, Δ) , then $\pi(\alpha) \propto \exp[\mathbb{E}(\log \mathcal{L}(E, \Delta|\alpha))]$. In our setting, the MDI prior for α is:

$$\pi(\alpha) \propto \exp \left[\log(\alpha) - \frac{1}{\alpha} \right]. \quad (6)$$

Using (5) and (6), the posterior density of α based on MDI prior is given by:

$$\pi(\alpha|E, \Delta) \propto \alpha^{\sum_{j=1}^{N_u} \delta_{[n-N_u+j]}} \exp \left[-\alpha \sum_{j=1}^{N_u} \log e_{j,u} \right] \alpha \exp \left[-\frac{1}{\alpha} \right],$$

and letting $a_u := \sum_{j=1}^{N_u} \log e_{j,u}$ and $b_u := \sum_{j=1}^{N_u} \delta_{[n-N_u+j]}$, the log-posterior density $\ell(\cdot|E, \Delta)$ of α satisfies:

$$\ell(\alpha|E, \Delta) \propto (1 + b_u) \log(\alpha) - a_u \alpha - \frac{1}{\alpha}.$$

From this, we can define the maximum a posteriori (or MAP) tail index estimator as:

$$\hat{\alpha}_{MAP}^{(MDI)} = \arg \max_{\alpha} \ell(\alpha|E, \Delta) := \frac{B + \sqrt{B^2 + 4A}}{2A}, \quad (7)$$

where $A = a_u$ and $B = 1 + b_u$. Letting $u = Z_{n,n-k}$ in (7), we obtain the following formal definition of our MAP estimator of α under MDI prior:

Definition 2.1. Let $k := k_n$ be a sequence such that $k \rightarrow \infty$ and $\frac{k}{n} \rightarrow 0$ as $n \rightarrow \infty$. Then the MAP estimator of the tail index α under MDI prior is given by:

$$\hat{\alpha}_{MAP}^{(MDI)} := \frac{B + \sqrt{B^2 + 4A}}{2A}, \quad (8)$$

where $A = \sum_{i=1}^k \log \left(\frac{Z_{n,n-k+i}}{Z_{n,n-k}} \right)$ and $B = 1 + \sum_{i=1}^k \delta_{[n-k+i]}$.

Remark 2.1. In order to simplify notations, the dependency of $\hat{\alpha}_{MAP}^{(MDI)}$ on k and n is not made explicit. This convention will also be adopted for the other estimators defined below.

2.2. Jeffreys prior

A second kind of non-informative prior is Jeffreys prior [19], which has an interesting invariant re-parametrization property. This prior is proportional to the square root of Fisher's information. If $\mathcal{L}(E, \Delta|\alpha)$ is the likelihood of a single observation (E, Δ) , Jeffreys prior can be written as:

$$\begin{aligned} \pi(\alpha) &\propto \left[-\mathbb{E} \left(\frac{\partial^2}{\partial \alpha^2} \log \mathcal{L}(E, \Delta|\alpha) \right) \right]^{\frac{1}{2}} \\ &\propto \frac{1}{\alpha}. \end{aligned} \quad (9)$$

Using (5) and (9), the posterior density of α based on Jeffreys prior is given by:

$$\pi(\alpha|E, \Delta) \propto \alpha^{(\sum_{j=1}^{N_u} \delta_{[n-N_u+j]})-1} \exp \left[-\alpha \sum_{j=1}^{N_u} \log e_{j,u} \right],$$

which coincides (up to some normalizing constants) with the probability density function of the Gamma distribution $\Gamma(\sum_{j=1}^{N_u} \delta_{[n-N_u+j]}, \sum_{j=1}^{N_u} \log e_{j,u})$. Based on this posterior, we construct two classical Bayesian estimators of α , namely the mean posterior estimator (MPE) and the maximum a posteriori (MAP) estimator.

- **mean posterior estimator (MPE):** this estimator is defined as the mean of the posterior distribution of α :

$$\hat{\alpha}_{MPE}^{(J)} := \frac{\sum_{j=1}^{N_u} \delta_{[n-N_u+j]}}{\sum_{j=1}^{N_u} \log e_{j,u}}. \quad (10)$$

Letting $u = Z_{n,n-k}$ in (10), we obtain the following formal definition of our MP estimator of the tail index α :

Definition 2.2. Let $k := k_n$ be a sequence such that $k \rightarrow \infty$ and $\frac{k}{n} \rightarrow 0$ as $n \rightarrow \infty$. Then the MP estimator of the tail index α under Jeffreys prior is given by:

$$\hat{\alpha}_{MPE}^{(J)} := \frac{\sum_{i=1}^k \delta_{[n-i+1]}}{\sum_{i=1}^k \log \left(\frac{Z_{n,n-i+1}}{Z_{n,n-k}} \right)}. \quad (11)$$

We note that $(\hat{\alpha}_{MPE}^{(J)})^{-1}$ coincides with the estimator of the EVI γ proposed by Einmahl [11] when censoring occurs.

- **maximum a posteriori estimator (MAP):** this estimator is defined as the mode of the posterior distribution of α :

$$\hat{\alpha}_{MAP}^{(J)} := \arg \max_{\alpha} \pi(\alpha|E, \Delta) = \frac{\sum_{j=1}^{N_u} \delta_{[n-N_u+j]} - 1}{\sum_{j=1}^{N_u} \log e_{j,u}}. \quad (12)$$

Letting $u = Z_{n,n-k}$ in (12), we obtain the following formal definition of our MAP estimator of the tail index α under Jeffreys prior:

Definition 2.3. Let $k := k_n$ be a sequence such that $k \rightarrow \infty$ and $\frac{k}{n} \rightarrow 0$ as $n \rightarrow \infty$. Then the MAP estimator of the tail index α under Jeffreys prior is given by:

$$\hat{\alpha}_{MAP}^{(J)} := \frac{\sum_{i=1}^k \delta_{[n-i+1]} - 1}{\sum_{i=1}^k \log \left(\frac{Z_{n,n-i+1}}{Z_{n,n-k}} \right)}. \quad (13)$$

2.3. Conjugate prior

In Subsections 2.1 and 2.2, prior elicitation is essentially data-driven since Jeffreys and MDI priors only take account of the information provided by the observations for estimating the tail index α . On the other hand, subjective priors (such as conjugate priors) attempt to bring prior knowledge about the phenomenon under study into the problem. Conjugate priors are particularly useful due to their mathematical convenience. Here, we propose a conjugate prior from the Gamma family $\Gamma(\cdot, \cdot)$.

If $\alpha \sim \Gamma(a, b)$ (with $a, b > 0$), we can easily obtain the posterior distribution of α given the observations (E, Δ) , as:

$$\alpha|(E, \Delta) \sim \Gamma \left(\sum_{j=1}^{N_u} \delta_{[n-N_u+j]} + a, \sum_{j=1}^{N_u} \log e_{j,u} + b \right).$$

Based on this, we propose two estimators of α under a conjugate Gamma prior.

Definition 2.4. Let $k := k_n$ be a sequence such that $k \rightarrow \infty$ and $\frac{k}{n} \rightarrow 0$ as $n \rightarrow \infty$. Let the prior density $\pi(\alpha)$ be a $\Gamma(a, b)$ distribution. Then:

1. the MP estimator of the tail index α is given by:

$$\hat{\alpha}_{MPE}^{(c)} := \frac{\sum_{i=1}^k \delta_{[n-i+1]} + a}{\sum_{i=1}^k \log \left(\frac{Z_{n,n-i+1}}{Z_{n,n-k}} \right) + b}, \quad (14)$$

2. the MAP estimator of the tail index α is given by:

$$\hat{\alpha}_{MAP}^{(c)} := \frac{\sum_{i=1}^k \delta_{[n-i+1]} + a - 1}{\sum_{i=1}^k \log \left(\frac{Z_{n,n-i+1}}{Z_{n,n-k}} \right) + b}. \quad (15)$$

2.4. Asymptotic properties of the Bayesian estimators

In this section, we investigate asymptotic properties of the proposed estimators. We first establish their strong consistency. Proof is given in Appendix A.

Theorem 2.1. *Let $k := k_n$ be a sequence such that $k \rightarrow \infty$ and $\frac{k}{n} \rightarrow 0$ as $n \rightarrow \infty$. Then the estimators $\hat{\alpha}_{MAP}^{(MDI)}$, $\hat{\alpha}_{MPE}^{(J)}$, $\hat{\alpha}_{MAP}^{(J)}$, $\hat{\alpha}_{MPE}^{(c)}$ and $\hat{\alpha}_{MAP}^{(c)}$ given by (8), (11), (13), (14) and (15) respectively converge almost surely to α as $n \rightarrow \infty$.*

Now, we discuss the limiting behavior of the posterior distribution of the proposed estimators. We need some further notations. Let

$$L_k(\alpha) = \log \alpha \sum_{i=1}^k \delta_{[n-k+i]} - \alpha \sum_{i=1}^k \log e_{[n-k+i]} - \sum_{i=1}^k \delta_{[n-k+i]} \log e_{[n-k+i]}$$

be the log-likelihood $\mathcal{L}(E, \Delta | \alpha)$ calculated with the threshold $u = Z_{n,n-k}$. Let $\mathcal{I}_k(\alpha) = -\frac{1}{k} \frac{\partial^2}{\partial \alpha^2} L_k(\alpha)$, $\mathcal{I}(\alpha) = \lim_{k \rightarrow \infty} \mathcal{I}_k(\alpha)$ and α_0 denote the "true" tail index value. We assume that α_0 belongs to an open interval $\mathcal{A} = (a_1, a_2)$ of \mathbb{R} , where $0 < a_1 < a_2 < \infty$. Finally, let $p := \mathbb{P}(X < Y)$ be the probability that X is uncensored and \xrightarrow{d} denote convergence in distribution. Then the following holds:

Theorem 2.2. *Let $k := k_n$ be a sequence such that $k \rightarrow \infty$ and $\frac{k}{n} \rightarrow 0$ as $n \rightarrow \infty$. Let $\pi(\alpha)$ be a prior density with finite expectation and such that $\pi(\cdot)$ is continuous and positive at α_0 . Let $\hat{\alpha}_\bullet$ be any of the estimators $\hat{\alpha}_{MAP}^{(MDI)}$, $\hat{\alpha}_{MPE}^{(J)}$, $\hat{\alpha}_{MAP}^{(J)}$, $\hat{\alpha}_{MPE}^{(c)}$, $\hat{\alpha}_{MAP}^{(c)}$. Then as $n \rightarrow \infty$,*

$$\sqrt{k}(\alpha_0 - \hat{\alpha}_\bullet) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{p\mathcal{I}(\alpha_0)}\right).$$

Proof is given in Appendix A.

3. Simulation study

In this section, we assess the finite-sample performance of the proposed estimators (8), (11), (13), (14) and (15) of the tail index of a heavy-tailed distribution under random censoring. More precisely, we assess influence of the censoring proportion and threshold choice on the behaviour of our estimators.

Study design. The simulation design is as follows. Let X and Y be independent random variables with cdf F and G respectively. X is the variable of interest and Y is the censoring random variable. We assume that F and G are heavy-tailed with tail index $\alpha > 0$ and $\beta > 0$ respectively. Let $p := \mathbb{P}(X < Y)$ be the probability that X is uncensored. Some straightforward calculations yield $\beta \approx \alpha \frac{1-p}{p}$ (e.g., [15]). Our simulation process is as follows:

1. We simulate a sample of n independent copies of (Z, δ) , where $Z = \min(X, Y)$, $\delta = \mathbb{1}_{\{X \leq Y\}}$ and the censoring proportion is $1 - p$. We consider various values for p , namely $p = 0.9, 0.75, 0.5, 0.25$ (that is, we allow the percentage of censoring in the right tail of X to be 10%, 25%, 50% and 75% respectively).
2. For each p , we compute the various proposed estimates, by incrementing the threshold (or fraction level) $\frac{k}{n}$ from $\frac{k_{min}}{n}$ to $\frac{k_{max}}{n}$. For notational simplicity, we use the following notations:

- $\hat{\alpha}_1(k)$ is the MPE estimator (11) in Definition 2.2,
- $\hat{\alpha}_2(k)$ is the MAP estimator (13) in Definition 2.3,
- $\hat{\alpha}_3(k)$ is the MAP estimator (8) in Definition 2.1,
- $\hat{\alpha}_4(k)$ is the MPE estimator (14) in Definition 2.4,
- $\hat{\alpha}_5(k)$ is the MAP estimator (15) in Definition 2.4.

3. Steps 1-2 are repeated m times, so that we obtain m realisations of each $\hat{\alpha}_\ell(k)$, for each p .
4. For each p , we calculate the empirical bias and RMSE (Root Mean Square Error) of each $\hat{\alpha}_\ell(k)$ (for $k = k_{min}, \dots, k_{max}$ and $\ell = 1, 2, 3, 4, 5$) over the m estimates.

We consider the following two simulation settings:

- X and Y are both distributed as Fréchet random variables, with cdf $F(x) = \exp(-x^{-\alpha})$ and $G(x) = \exp(-x^{-\beta})$ respectively,
- X and Y are both distributed as Generalized Pareto Distribution (GPD) with cdf $F(x) = 1 - (1 + \frac{1}{\alpha}x)^{-\alpha}$ and $G(x) = 1 - (1 + \frac{1}{\beta}x)^{-\beta}$ respectively.

In either case, we take $\alpha = 0.5$ and $\alpha = 0.75$ and for each $p = 0.9, 0.75, 0.5, 0.25$, we take $\beta = \alpha \frac{1-p}{p}$. The simulations are conducted using the statistical software R [24]. Results are provided for a sample size $n = 1000$ and $m = 1000$ simulated samples, with $k_{min} = 20$ and $k_{max} = 250$.

Results for Fréchet model. For each $\ell = 1, \dots, 5$, we plot the empirical bias and RMSE of $\hat{\alpha}_\ell$ as functions of k for $p = 0.9, 0.75, 0.5, 0.25$ (Figure 1: $\alpha = 0.5$ and Figure 2 : $\alpha = 0.75$). From these figures, it appears as expected that the bias and RMSE of the proposed estimators decrease when the censoring proportion decreases. The MAP estimator under MDI prior $\hat{\alpha}_3$ (see (8), Definition 2.1) might be regarded as the best among all estimators $\hat{\alpha}_\ell$. In particular, this estimator outperforms all others in terms of RMSE for almost every k .

In Tables 1 and 2, we report the averaged (over the m simulated samples) value, empirical bias and RMSE of each $\hat{\alpha}_\ell$ at the optimal fraction level $k_{opt} = \arg \min_k RMSE[\hat{\alpha}_\ell(k)]$ (Table 1: $\alpha = 0.5$ and Table 2: $\alpha = 0.75$). These tables confirm the superiority of the MAP estimator under MDI prior $\hat{\alpha}_3$. In particular, at the optimal fraction level k_{opt} , $\hat{\alpha}_3$ appears to be less biased than all other estimators, for every p .

Results for the GPD model. Results for the GPD are given in Appendix B. The conclusions are similar to conclusions for Fréchet model.

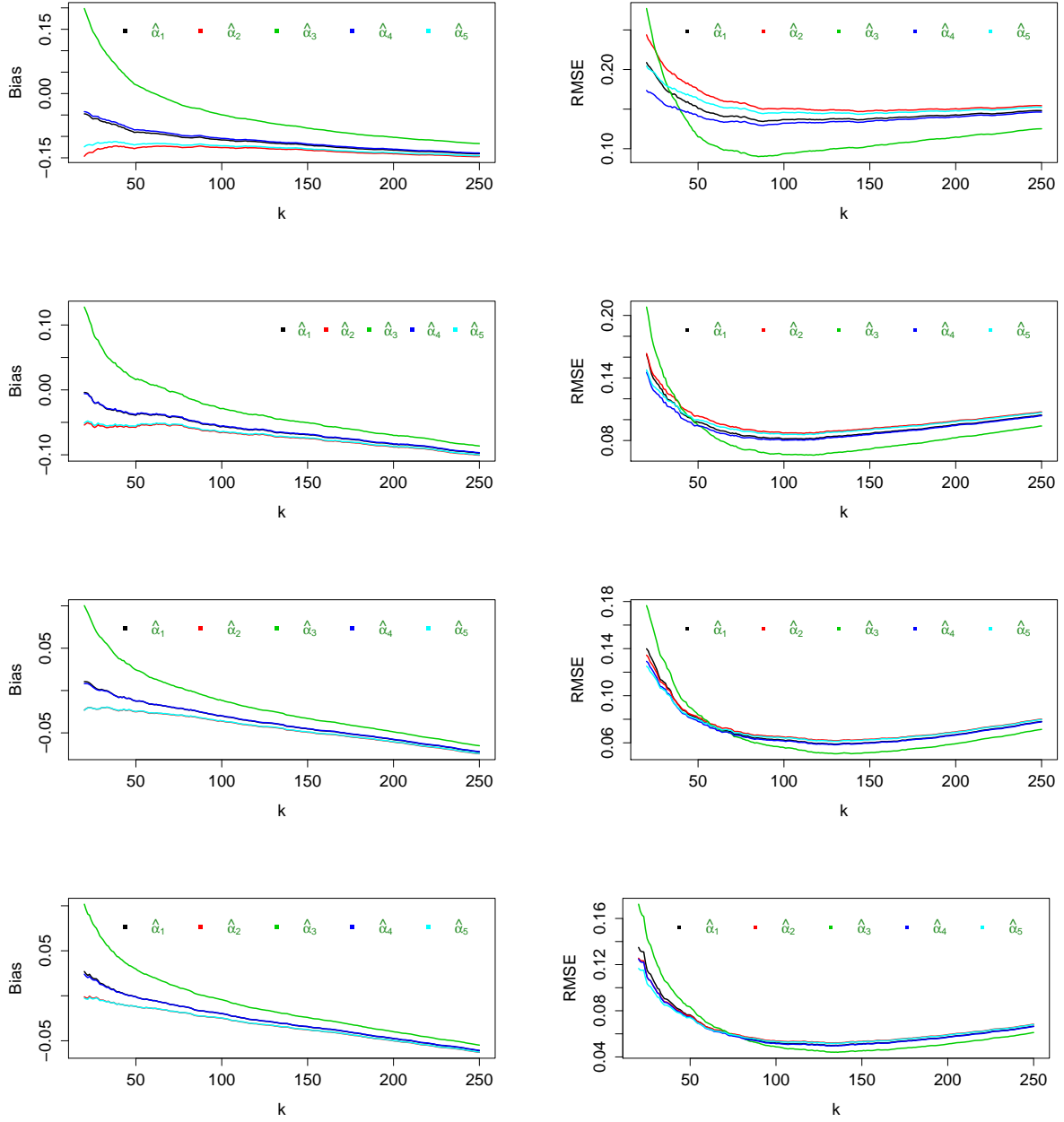


Figure 1: Empirical bias (left) and RMSE (right) of $\hat{\alpha}_l$ ($l = 1, \dots, 5$) for $\alpha = 0.5$ and $p = 0.25$ (1st line:), $p = 0.5$ (2nd line), $p = 0.75$ (3rd line) and $p = 0.9$ (4th line).

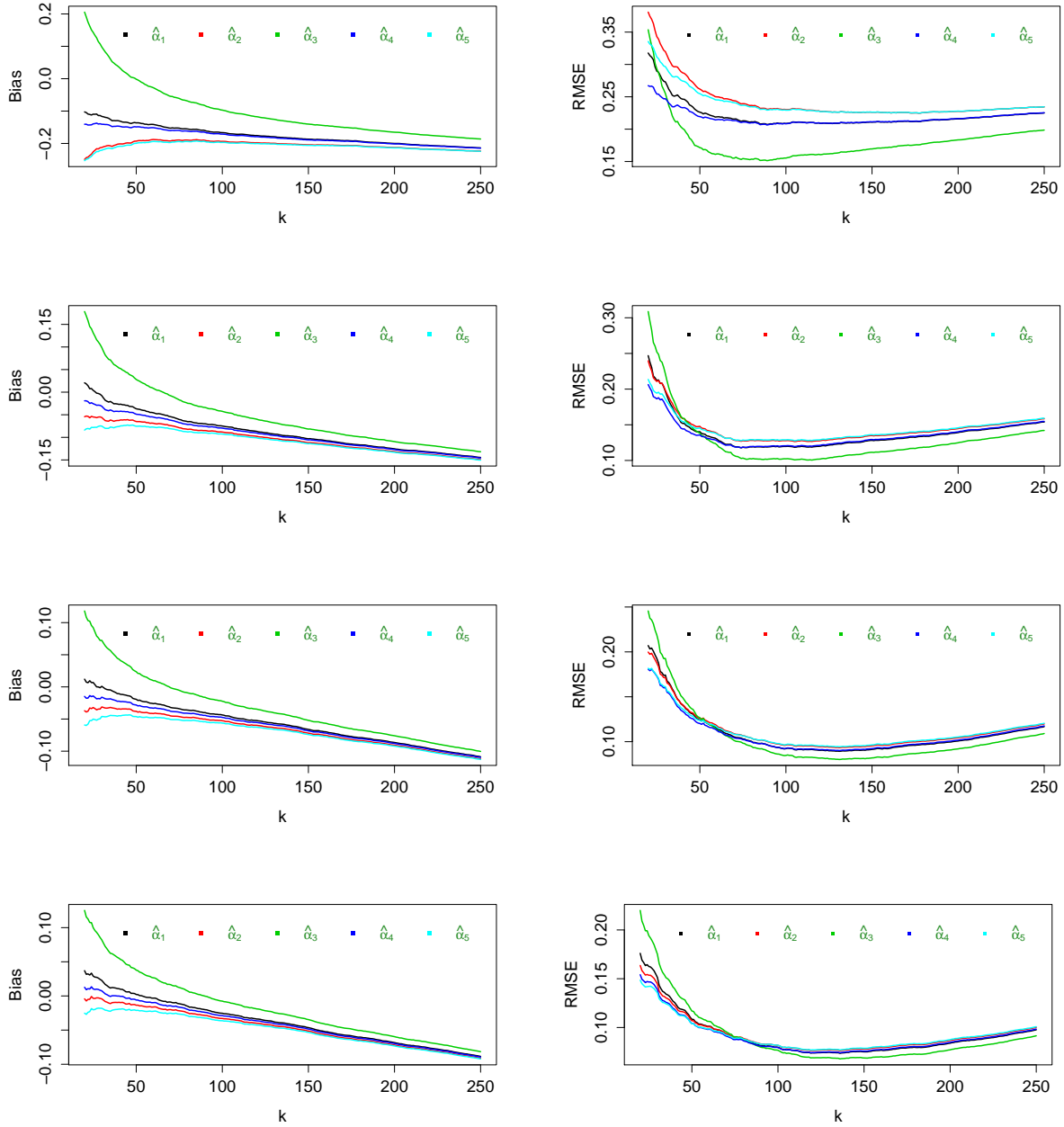


Figure 2: Empirical bias (left) and RMSE (right) of $\hat{\alpha}_\ell$ ($\ell = 1, \dots, 5$) for $\alpha = 0.75$ and $p = 0.25$ (1st line:), $p = 0.5$ (2nd line), $p = 0.75$ (3rd line) and $p = 0.9$ (4th line).

4. Randomizing the fraction level

In the previous section, the $(n-k)$ -th order statistic $Z_{n,n-k}$ was used as threshold in the excesses $e_{j,u}$ and k (or equivalently, the fraction level k/n) was chosen so as to minimize the RMSE. Obviously, this procedure cannot be applied in a practical data analysis since the true tail index α used to calculate the

		k_{opt}	Estimator	RMSE	Bias
p=0.9	$\hat{\alpha}_1$	133	0.4701402	0.04985720	-0.02985977
	$\hat{\alpha}_2$	133	0.4663334	0.05199783	-0.03366664
	$\hat{\alpha}_3$	137	0.4804375	0.04398932	-0.01956245
	$\hat{\alpha}_4$	133	0.4703423	0.04950047	-0.02965768
	$\hat{\alpha}_5$	133	0.4665644	0.05162741	-0.03343560
p=0.75	$\hat{\alpha}_1$	131	0.4610741	0.05884770	-0.03892594
	$\hat{\alpha}_2$	131	0.4565036	0.06172861	-0.04349636
	$\hat{\alpha}_3$	131	0.4752784	0.05071124	-0.02472156
	$\hat{\alpha}_4$	130	0.4616714	0.05834371	-0.03832862
	$\hat{\alpha}_5$	130	0.4571051	0.06120118	-0.04289495
p=0.5	$\hat{\alpha}_1$	113	0.4428444	0.08150309	-0.05715557
	$\hat{\alpha}_2$	113	0.4349585	0.08687673	-0.06504148
	$\hat{\alpha}_3$	118	0.4639250	0.06592663	-0.03607495
	$\hat{\alpha}_4$	111	0.4480398	0.08020518	-0.05196024
	$\hat{\alpha}_5$	113	0.4359118	0.08560104	-0.06408818
p=0.25	$\hat{\alpha}_1$	89	0.3979008	0.1343862	-0.10209924
	$\hat{\alpha}_2$	144	0.3704152	0.1468569	-0.12958479
	$\hat{\alpha}_3$	86	0.4656990	0.0900378	-0.03430104
	$\hat{\alpha}_4$	89	0.4017657	0.1292944	-0.09823434
	$\hat{\alpha}_5$	144	0.3734453	0.1434458	-0.12655470

Table 1: Optimal results for Fréchet model with $n = 1000$, $m = 1000$ and $\alpha = 0.5$.

		k_{opt}	Estimator	RMSE	Bias
p=0.9	$\hat{\alpha}_1$	137	0.7102997	0.07354345	-0.03970025
	$\hat{\alpha}_2$	137	0.7048372	0.07626555	-0.04516285
	$\hat{\alpha}_3$	137	0.7233166	0.06780636	-0.02668343
	$\hat{\alpha}_4$	137	0.7079724	0.07410573	-0.04202759
	$\hat{\alpha}_5$	120	0.7076827	0.07689530	-0.04231728
p=0.75	$\hat{\alpha}_1$	132	0.6930671	0.08929557	-0.05693289
	$\hat{\alpha}_2$	132	0.6863281	0.09336546	-0.06367190
	$\hat{\alpha}_3$	132	0.7093273	0.08017017	-0.04067273
	$\hat{\alpha}_4$	132	0.6904320	0.09015946	-0.05956802
	$\hat{\alpha}_5$	132	0.6837832	0.09433285	-0.06621679
p=0.5	$\hat{\alpha}_1$	74	0.6909873	0.1176917	-0.05901272
	$\hat{\alpha}_2$	105	0.6599687	0.1265239	-0.09003127
	$\hat{\alpha}_3$	115	0.6940075	0.1005289	-0.05599250
	$\hat{\alpha}_4$	74	0.6839079	0.1178655	-0.06609215
	$\hat{\alpha}_5$	105	0.6558685	0.1277539	-0.09413154
p=0.25	$\hat{\alpha}_1$	89	0.5910295	0.2071481	-0.15897046
	$\hat{\alpha}_2$	177	0.5438101	0.2246926	-0.20618988
	$\hat{\alpha}_3$	89	0.6687678	0.1513115	-0.08123217
	$\hat{\alpha}_4$	89	0.5854156	0.2064865	-0.16458441
	$\hat{\alpha}_5$	177	0.5424875	0.2248853	-0.20751253

Table 2: Optimal results for Fréchet model with $n = 1000$, $m = 1000$ and $\alpha = 0.75$.

RMSE is unknown. Thus, in this section, we develop an alternative approach for tackling this threshold issue.

We propose to consider the fraction level $t := k/n$ as a random quantity. As k/n lies in $]0, 1[$, we suggest to use a Beta distribution for t , with probability density function:

$$\pi_u(t) = \frac{1}{B(a, b)} \frac{(t - t_{\min})^{a-1} (t_{\max} - t)^{b-1}}{(t_{\max} - t_{\min})^{a+b-1}} \mathbb{1}_{\{t_{\min} \leq t \leq t_{\max}\}}, \quad (16)$$

where $B(a, b) := \int_0^1 s^{a-1} (1-s)^{b-1} ds$ ($a > 0, b > 0$) is the Beta function and t_{\min} and t_{\max} are suitable bounds for the fraction level. Then we propose the following Monte Carlo approach for inferring on α from a sample of observations $(Z_i, \delta_i)_{i=1, \dots, n}$.

First, we simulate $x \sim \text{Beta}(a, b)$ and we calculate $t_x = (t_{\max} - t_{\min})x + t_{\min}$. Then we obtain $k_x = \lfloor nt_x \rfloor$ (where $\lfloor \cdot \rfloor$ denotes the integer part) and we calculate the estimate $\hat{\alpha}_\ell$ with $k = k_x$. This procedure is repeated N times, which yields N Monte Carlo realisations $\hat{\alpha}_\ell^{(1)}, \dots, \hat{\alpha}_\ell^{(N)}$ of $\hat{\alpha}_\ell$. Finally, α can be estimated from these N realisations (by taking their empirical mean, median or mode, for example).

We illustrate this procedure in a short simulation study. A sample of size $n = 1000$ of (Z, δ) is obtained by simulating n observations of X and Y from Pareto distributions with tail index $\alpha = 2$ and $\beta = \alpha \frac{1-p}{p}$ respectively, where $p = 0.25, 0.5, 0.75$ and 0.9 (that is, the censoring fraction is 75%, 50%, 25% and 10% respectively). Based on results of Section 3, we restrict our attention to objective priors. Hence, the proposed randomization procedure is applied to the MP estimator of α under Jeffreys prior (*i.e.*, $\hat{\alpha}_1$) and to the MAP estimator under Jeffreys and MDI priors ($\hat{\alpha}_2$ and $\hat{\alpha}_3$ respectively). We take $t_{\min} = 0.01$ and $t_{\max} = 0.4$. Note that under these values, k is allowed to range from 10 to 400, with a higher concentration of values around $k = 300$. This is a reasonable choice in view of the sample size. We also take $a = 4$ and $b = 1.5$ (a procedure for choosing a and b will be discussed in Section 5). Finally, we let $N = 10000$.

For each $p \in \{0.25, 0.5, 0.75, 0.9\}$ and $l \in \{1, 2, 3\}$, we obtain the empirical mean, median and mode of the N Monte Carlo estimates $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$ (see Table 3) and we plot their histogram and estimated density (see Figure 3).

		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
p=0.25	Mode	2.037	1.983	2.048
	Median	2.074	2.015	2.085
	Mean	2.052	2.014	2.096
p=0.50	Mode	2.014	1.992	2.027
	Median	2.024	2.008	2.034
	Mean	2.002	1.988	2.027
p=0.75	Mode	1.985	1.975	2.033
	Median	2.026	2.015	2.044
	Mean	2.038	2.027	2.054
p=0.90	Mode	1.973	1.938	1.997
	Median	2.044	1.995	2.051
	Mean	2.007	1.998	2.057

Table 3: Randomization of the fraction level: empirical mean, median and mode of the Monte Carlo estimates $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$ of the tail index $\alpha = 2$.

From Figure 3, histograms are concentrated around the true tail index value $\alpha = 2$ (represented by the vertical dotted line) in every case. Moreover, the empirical mean, median and mode of the $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$ appear to provide satisfactory approximations of α , for every p and $\ell = 1, 2, 3$.

Overall, combining the proposed Bayesian estimators with the randomization tool described above seems to provide a relevant approach for estimating α . This approach is now illustrated on a real data set.

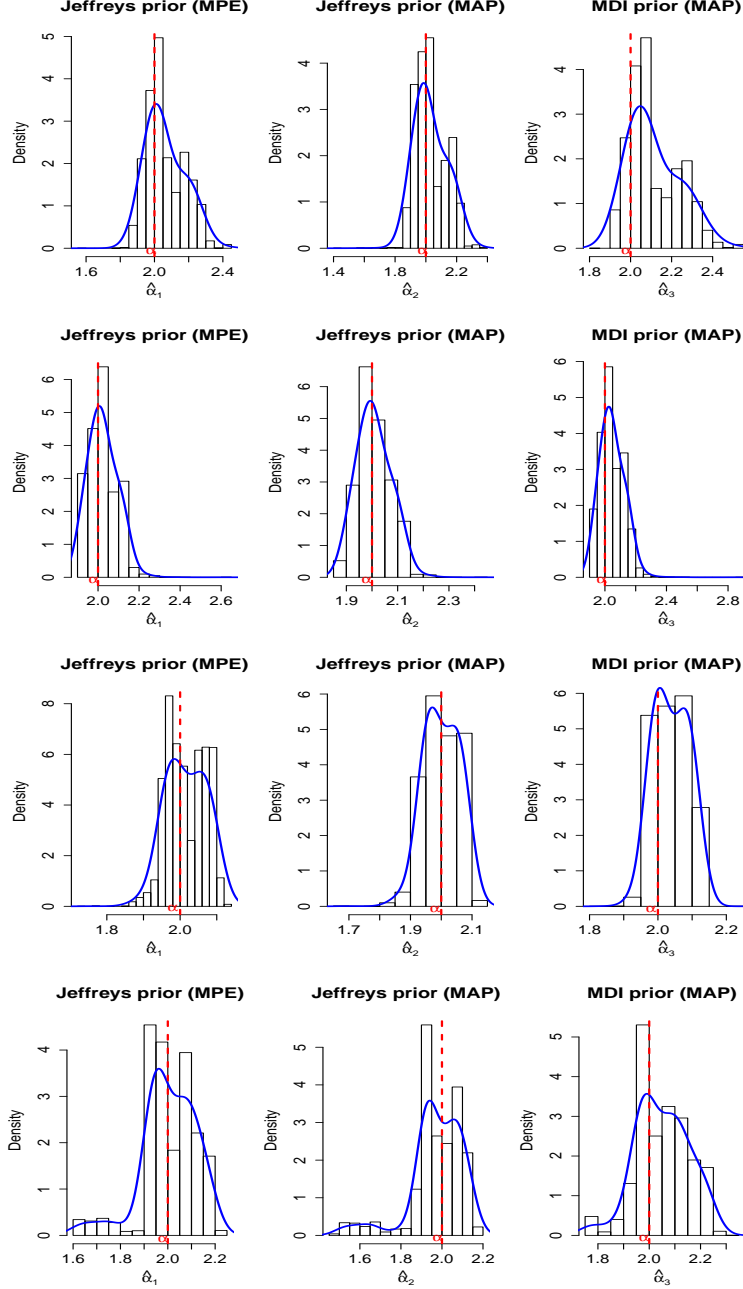


Figure 3: Randomization of the fraction level: histograms of the Monte Carlo estimates $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$ of the tail index $\alpha = 2$ for $p = 0.25$ (1st line), $p = 0.5$ (2nd line), $p = 0.75$ (3rd line), $p = 0.9$ (4th line).

5. An application to AIDS survival data

In this section, we illustrate our methodology on a set of AIDS survival data. The dataset contains $n = 2843$ patients diagnosed with AIDS in Australia before 1 July 1991. The source of the data is Dr P.J. Solomon and the Australian National Centre in HIV Epidemiology and Clinical Research (see [29]). The data are available in the R package "MASS" [24]. Information on each patient includes gender, date of diagnosis, age at diagnosis, date of death or end of observation and an indicator which equals 1 if the patient died and 0 otherwise. 1761 patients died. The other survival times are right-censored. Our

objective is to estimate the tail index α of the cdf F of the survival time. We calculate our estimators on the whole dataset of patients, without distinction of sex or age.

Considering only objective priors for α , we calculate the MP and MAP estimators under Jeffreys prior ($\hat{\alpha}_1$ and $\hat{\alpha}_2$ respectively) and the MAP estimator under MDI prior ($\hat{\alpha}_3$). To choose the fraction level k/n , we first rely on a method proposed by Einmahl *et al.* (2008) [11]. Then we use the randomizing approach described in Section 4. Einmahl *et al.* (2008) [11] propose to plot the proportion $\hat{p}(k) = \frac{1}{k} \sum_{i=1}^n \delta_i \mathbb{1}_{\{Z_i > Z_{n,n-k}\}}$ of non-censored observations in the k largest Z_i 's as a function of k (see Figure (4)).

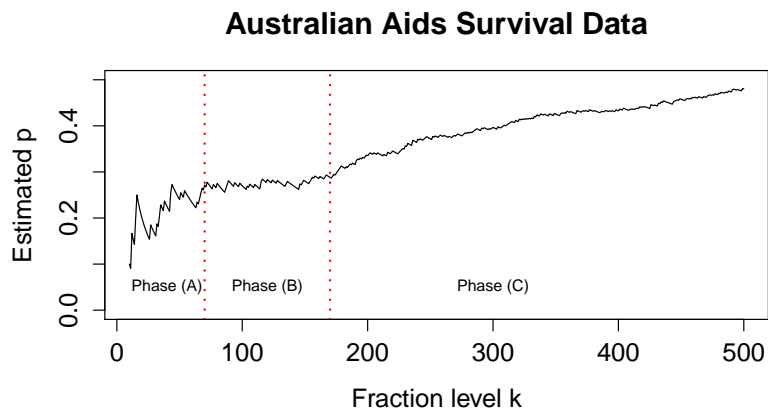


Figure 4: Proportion of uncensored Z_i 's among the k largest observations.

Three distinct phases can be distinguished on Figure (4). In phase (A), the behaviour of $k \mapsto \hat{p}(k)$ is somewhat erratic. Then, there is stable part (phase (B), where k ranges approximately from 70 to 170 and the average $\hat{p}(k)$ is approximately 0.261) and an increasing part (phase (C)). Einmahl *et al.* (2008) [11] suggest to estimate α based on the k -values of phase (B). Indeed, the estimate of α should be quite stable within this range of k . On Figure (5), we plot the estimates $\hat{\alpha}_1(k)$, $\hat{\alpha}_2(k)$ and $\hat{\alpha}_3(k)$ as functions of k .

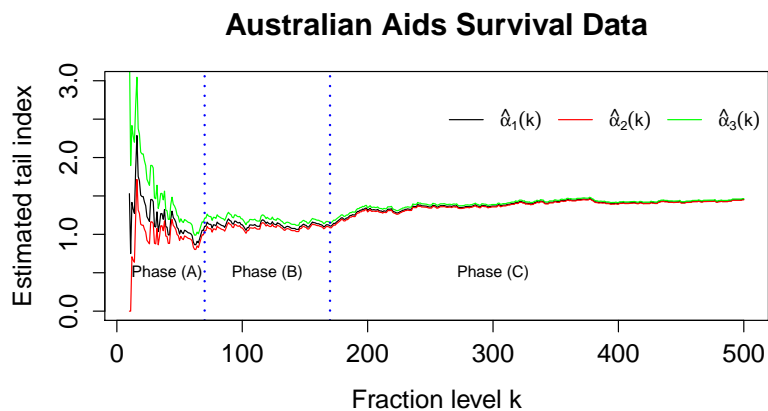


Figure 5: Tail index estimation for Australian AIDS survival data.

We clearly distinguish the same three phases (A), (B) and (C) as in Figure (4). All three estimates appear to be close to each other and quite stable in phase (B). The average value of $\hat{\alpha}_\ell(k)$ on phase (B)

is approximately 1.26 (for $\ell = 1, 2, 3$). Based on Einmahl *et al.* [11] empirical methodology, one may thus retain 1.26 as an estimate of α .

We now apply our randomization procedure. We take $t_{\min} = 0.01$ and $t_{\max} = 0.4$, so that k ranges from 28 to 1137 (this range includes and goes far beyond the stable phase (B)). We consider several values for the parameters $a, b > 0$ of the Gamma distribution used to randomize the fraction level, namely $(a, b) \in \mathcal{S} := \{1, 1.5, 2, 2.5, 3, 3.5, 4\} \times \{1, 1.5, 2, 2.5, 3, 3.5, 4\}$. For every $(a, b) \in \mathcal{S}$, we apply the proposed randomization procedure with $N = 10000$ and we calculate the standard deviation (Std.) of the resulting N Monte Carlo estimates $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$. Then, we retain the value (a^*, b^*) such that the standard deviation is minimum. Finally, we propose to infer on α from the $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$ obtained with $(a, b) = (a^*, b^*)$.

Table 4 reports standard deviations for $\hat{\alpha}_1$. The minimum is achieved when $(a, b) = (4, 3)$ (Std. = 0.117). Standard deviations for the pairs $(4, 2)$ (Std. = 0.120) and $(4, 2.5)$ (Std. = 0.118) are close to the minimum, thus we also retain these values for making inference on α . The minimum standard deviations for $\hat{\alpha}_2$ and $\hat{\alpha}_3$ are achieved for the same values of (a, b) . The corresponding tables are omitted for conciseness.

		b						
		1.0	1.5	2.0	2.5	3.0	3.5	4.0
α	1.0	0.353	0.381	0.410	0.443	0.456	0.463	0.469
	1.5	0.243	0.282	0.315	0.342	0.372	0.395	0.409
	2.0	0.194	0.202	0.233	0.263	0.286	0.312	0.341
	2.5	0.156	0.166	0.175	0.194	0.121	0.244	0.263
	3.0	0.134	0.141	0.144	0.152	0.167	0.187	0.205
	3.5	0.127	0.130	0.127	0.128	0.134	0.141	0.159
	4.0	0.122	0.123	0.120	0.118	0.117	0.124	0.127

Table 4: Standard deviations of the Monte Carlo estimates $\hat{\alpha}_1^{(j)}, j = 1, \dots, N$ for various values of (a, b) . The minimum standard deviations are indicated in bold.

We can now infer on α from the $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$ obtained with (a^*, b^*) (by taking their empirical mean, median or mode, for example). We report some summary statistics for $\hat{\alpha}_1, \hat{\alpha}_2$ and $\hat{\alpha}_3$ when $(a^*, b^*) = (4, 2), (4, 2.5), (4, 3)$ (see Table 5). We also plot the corresponding histograms of the $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$ (see Figure 6).

		Std.	Min.	1st Qu.	Mode	Median	Mean	3rd Qu.	Max.
(*)	$\hat{\alpha}_1$	0.120	1.112	1.272	1.293	1.390	1.381	1.470	2.602
	$\hat{\alpha}_2$	0.119	1.111	1.268	1.287	1.386	1.378	1.466	2.570
	$\hat{\alpha}_3$	0.121	1.116	1.277	1.298	1.398	1.388	1.478	2.645
(**)	$\hat{\alpha}_1$	0.118	1.119	1.306	1.327	1.398	1.402	1.487	2.614
	$\hat{\alpha}_2$	0.117	1.118	1.303	1.436	1.395	1.398	1.483	2.583
	$\hat{\alpha}_3$	0.119	1.122	1.314	1.307	1.405	1.410	1.495	2.657
(***)	$\hat{\alpha}_1$	0.117	1.124	1.323	1.361	1.425	1.419	1.506	2.589
	$\hat{\alpha}_2$	0.117	1.123	1.317	1.458	1.422	1.414	1.502	2.557
	$\hat{\alpha}_3$	0.119	1.128	1.335	1.471	1.435	1.427	1.511	2.633

Table 5: Summary statistics for the Monte Carlo estimates $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$ obtained with $(a, b) = (4, 2)$ (see (*)), $(a, b) = (4, 2.5)$ (see (**)) and $(a, b) = (4, 3)$ (see (***)).

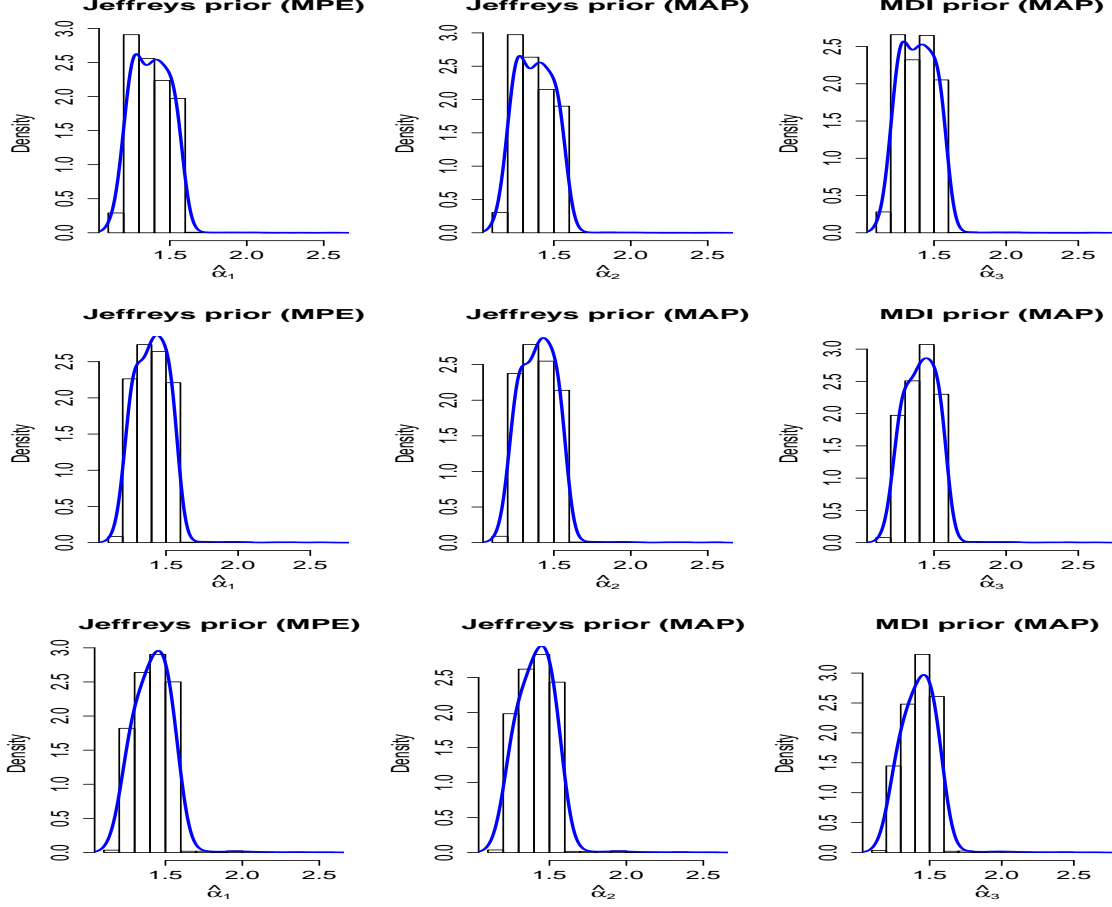


Figure 6: Randomization of the fraction level: histograms of the $\hat{\alpha}_\ell$ ($\ell = 1, 2, 3$) with $(a^*, b^*) = (4, 2)$ (1st line), $(a^*, b^*) = (4, 2.5)$ (2nd line), $(a^*, b^*) = (4, 3)$ (3rd line).

From Figure 6, histograms are quite symmetric and show limited spread. Thus, it seems reasonable to estimate α by any of the usual measures of central tendency. Moreover, from Table 5, the empirical means and medians of the $\hat{\alpha}_\ell$ ($\ell = 1, 2, 3$) are close to each other, for every $(a, b) = (4, 2), (4, 2.5), (4, 3)$. From these results, one may retain 1.4 as a reasonable and consensual (over the $\hat{\alpha}_\ell, \ell = 1, 2, 3$) approximation of α .

6. Conclusion and perspectives

In this paper, we address estimation of the tail index of a heavy-tailed distribution when data are randomly right-censored. We constructed two kinds of Bayesian estimators, namely mean posterior and maximum a posteriori estimators, for various objective and subjective prior distributions. Our simulation results indicate better performance of estimators based on objective MDI and Jeffreys priors. We also proposed an original automatic procedure for selecting the threshold (or fraction level k/n) required to estimate the tail index. The whole procedure was applied to a set of real data and provided coherent results.

Now, several issues still deserve attention. First, we mentioned in introduction that estimation of the tail index constitutes usually a first step in an extreme value analysis. Extreme quantile estimation often comes as a second step. Extreme quantiles of the distribution of X are quantities of the form

$$F^{\leftarrow}(1 - p) = \inf\{y : F(y) \geq 1 - p\},$$

where p is so small that this quantile falls beyond the range of the observed data X_1, \dots, X_n . Estimation of extreme quantiles has become a crucial issue in many domains (*e.g.*, determination of the value at risk in financial risk management, determination of the return period of extreme precipitation in meteorology...). We are currently working on Bayesian estimation of extreme quantiles from censored data.

In practice, it often arises that some covariate information W is available to the investigator and the distribution of X depends on W . In this case, interest turn to estimation of the conditional tail index $\alpha(w)$. Adapting our Bayesian approach to this setting also constitutes a topic for our future research.

Appendix A. Proofs.

In this appendix, we outline the proofs of Theorem 2.1 and Theorem 2.2.

Proof of Theorem 2.1.

We prove strong consistency of $\hat{\alpha}_{MAP}^{(MDI)}$. First, it follows from [9] (Theorem 2.3) and [11] that $\frac{1}{k} \sum_{i=1}^k \delta_{[n-i+1]}$ converges almost surely (a.s.) to $\frac{\alpha}{\alpha+\beta}$ as $n \rightarrow \infty$. From [7], $\frac{1}{k} \sum_{i=1}^k \log \left(\frac{Z_{n,n-i+1}}{Z_{n,n-k}} \right)$ converges a.s. to the EVI $\frac{1}{\alpha+\beta}$ of Z , as $n \rightarrow \infty$. By the continuous mapping theorem (see [28], for example), $\hat{\alpha}_{MAP}^{(MDI)}$ converges a.s. to $\frac{\alpha}{\alpha+\beta} \cdot (\alpha + \beta) = \alpha$ as $n \rightarrow \infty$. Proofs for $\hat{\alpha}_{MPE}^{(J)}$, $\hat{\alpha}_{MAP}^{(J)}$, $\hat{\alpha}_{MPE}^{(c)}$ and $\hat{\alpha}_{MAP}^{(c)}$ are similar and are thus omitted. \square

Proof of Theorem 2.2.

We first establish asymptotic normality of $\hat{\alpha}_{MPE}^{(J)}$. Proof proceeds along the same lines as proof of Theorem 4.2 in [14], which establishes asymptotic normality of the posterior distribution of a strongly consistent solution of a likelihood equation (note that $\hat{\alpha}_{MPE}^{(J)}$ coincides with the solution of the likelihood equation $\frac{\partial L_k(\alpha)}{\partial \alpha} = 0$ and is strongly consistent by Theorem 2.1). Some regularity conditions are needed.

First, we note that the support of the density of (Z, δ) is the same for all $\alpha \in \mathcal{A}$. Moreover, $L_1(\alpha)$ is thrice differentiable with respect to α and if $U = (\alpha_0 - \delta, \alpha_0 + \delta)$ is a neighborhood of α_0 , we have $\sup_{\alpha \in U} \left| \frac{\partial^3 L_1(\alpha)}{\partial \alpha^3} \right| < M(Z, \delta)$, where $\mathbb{E}M(Z, \delta) < \infty$. Finally, for any $\delta > 0$, there exists an $\epsilon > 0$ such that for all sufficiently large k , we have

$$\sup_{|\alpha - \hat{\alpha}_{MPE}^{(J)}| > \delta} \frac{1}{k} \left[L_k(\alpha) - L_k(\hat{\alpha}_{MPE}^{(J)}) \right] < -\epsilon. \quad (17)$$

To see this, consider the Taylor expansion

$$L_k(\alpha) = L_k(\hat{\alpha}_{MPE}^{(J)}) + (\alpha - \hat{\alpha}_{MPE}^{(J)}) \frac{\partial L_k(\hat{\alpha}_{MPE}^{(J)})}{\partial \alpha} + \frac{(\alpha - \hat{\alpha}_{MPE}^{(J)})^2}{2} \frac{\partial^2 L_k(\tilde{\alpha})}{\partial \alpha^2},$$

where $\tilde{\alpha}$ belongs to the line segment between α and $\hat{\alpha}_{MPE}^{(J)}$. Then

$$\frac{1}{k} \left[L_k(\alpha) - L_k(\hat{\alpha}_{MPE}^{(J)}) \right] = -\frac{(\alpha - \hat{\alpha}_{MPE}^{(J)})^2}{2\tilde{\alpha}^2} \frac{1}{k} \sum_{i=1}^k \delta_{[n-k+i]}.$$

Let $\xi > 0$ be arbitrary small. For k sufficiently large, $\frac{1}{k} \sum_{i=1}^k \delta_{[n-k+i]} < p + \xi$ and thus, for all α such that $|\alpha - \hat{\alpha}_{MPE}^{(J)}| > \delta$,

$$\frac{1}{k} \left[L_k(\alpha) - L_k(\hat{\alpha}_{MPE}^{(J)}) \right] < -\frac{\delta^2}{2a_2^2} (p + \xi).$$

Finally, (17) follows by letting $\epsilon := \frac{\delta^2}{2a_2^2} (p + \xi)$. We now establish asymptotic normality of $W_k := \sqrt{k}(\alpha_0 - \hat{\alpha}_{MPE}^{(J)})$.

The posterior density of W_k can be written as

$$\pi_{W_k}(w|E, \Delta) := C_k^{-1} \pi(\hat{\alpha}_{MPE}^{(J)} + k^{-1/2}w) \exp \left[L_k(\hat{\alpha}_{MPE}^{(J)} + k^{-1/2}w) - L_k(\hat{\alpha}_{MPE}^{(J)}) \right],$$

where $C_k = \int_{\mathbb{R}} \pi(\hat{\alpha}_{MPE}^{(J)} + k^{-1/2}w) \exp[L_k(\hat{\alpha}_{MPE}^{(J)} + k^{-1/2}w) - L_k(\hat{\alpha}_{MPE}^{(J)})] dw$. We show that $\pi_{W_k}(w|E, \Delta)$ converges in L^1 to the probability density function of the random variable $\mathcal{N}(0, \frac{1}{p\mathcal{I}(\alpha_0)})$, that is, we show that

$$I_k := \int_{\mathbb{R}} \left| \pi_{W_k}(w|E, \Delta) - \sqrt{\frac{p\mathcal{I}(\alpha_0)}{2\pi}} \exp \left[-\frac{w^2}{2} p\mathcal{I}(\alpha_0) \right] \right| dw \rightarrow 0 \text{ as } k \rightarrow \infty.$$

Convergence in L^1 -norm will imply that W_k converges in distribution to $\mathcal{N}(0, \frac{1}{p\mathcal{I}(\alpha_0)})$. Let

$$g_k(w) = \pi(\hat{\alpha}_{MPE}^{(J)} + k^{-1/2}w) \exp \left[L_k(\hat{\alpha}_{MPE}^{(J)} + k^{-1/2}w) - L_k(\hat{\alpha}_{MPE}^{(J)}) \right] - \pi(\alpha_0) \exp \left[-\frac{w^2}{2} p\mathcal{I}(\alpha_0) \right].$$

Then we have:

$$I_k \leq C_k^{-1} \int_{\mathbb{R}} |g_k(w)| dw + \int_{\mathbb{R}} \left| C_k^{-1} \pi(\alpha_0) - \sqrt{\frac{p\mathcal{I}(\alpha_0)}{2\pi}} \exp \left[-\frac{w^2}{2} p\mathcal{I}(\alpha_0) \right] \right| dw. \quad (18)$$

We first prove that $\int_{\mathbb{R}} |g_k(w)| dw \rightarrow 0$ as $n \rightarrow \infty$ (and thus, as $k \rightarrow \infty$). Let $\mathbb{R} = R_1 \cup R_2$, where $R_1 = \{w : |w| > k^{1/2}\delta\}$ and $R_2 = \{w : |w| < k^{1/2}\delta\}$ for some $\delta > 0$. We show that $\int_{R_j} |g_k(w)| dw \rightarrow 0$ as $n \rightarrow \infty$, for $j = 1, 2$. Note first that

$$\begin{aligned} \int_{R_1} |g_k(w)| dw &\leq \int_{R_1} \pi(\hat{\alpha}_{MPE}^{(J)} + k^{-1/2}w) \exp \left[L_k(\hat{\alpha}_{MPE}^{(J)} + k^{-1/2}w) - L_k(\hat{\alpha}_{MPE}^{(J)}) \right] dw \\ &\quad + \int_{R_1} \pi(\alpha_0) \exp \left[-\frac{w^2}{2} p\mathcal{I}(\alpha_0) \right] dw \\ &\leq \exp(-k\epsilon) + \pi(\alpha_0) \int_{R_1} \exp \left[-\frac{w^2}{2} p\mathcal{I}(\alpha_0) \right] dw, \end{aligned}$$

where the second inequality follows from (17). The term $\int_{R_1} \exp \left[-\frac{w^2}{2} p\mathcal{I}(\alpha_0) \right] dw$ is proportional to $\Psi(-k^{1/2}\delta)$, where Ψ is the cdf of the normal distribution $\mathcal{N}(0, \frac{1}{p\mathcal{I}(\alpha_0)})$. Therefore, this term converges to 0 as $n \rightarrow \infty$ and finally, $\int_{R_1} |g_k(w)| dw \rightarrow 0$ as $n \rightarrow \infty$. The proof that $\int_{R_2} |g_k(w)| dw \rightarrow 0$ as $n \rightarrow \infty$ is similar and is therefore omitted.

Now, we consider the second term on the right-hand side of (18). Note first that

$$\left| C_k - \pi(\alpha_0) \sqrt{\frac{2\pi}{p\mathcal{I}(\alpha_0)}} \right| \leq \int_{\mathbb{R}} |g_k(w)| dw.$$

This implies that $C_k^{-1} \pi(\alpha_0) \rightarrow \sqrt{\frac{p\mathcal{I}(\alpha_0)}{2\pi}}$ as $n \rightarrow \infty$ and thus, $\int_{\mathbb{R}} |C_k^{-1} \pi(\alpha_0) - \sqrt{\frac{p\mathcal{I}(\alpha_0)}{2\pi}}| \exp \left[-\frac{w^2}{2} p\mathcal{I}(\alpha_0) \right] dw \rightarrow 0$ as $n \rightarrow \infty$.

Finally, $I_k \rightarrow 0$ as $n \rightarrow \infty$ and thus, $W_k := \sqrt{k}(\alpha - \hat{\alpha}_{MPE}^{(J)})$ converges in distribution to $\mathcal{N}(0, \frac{1}{p\mathcal{I}(\alpha_0)})$.

We now establish asymptotic normality of $\hat{\alpha}_{MAP}^{(J)}, \hat{\alpha}_{MAP}^{(MDI)}, \hat{\alpha}_{MPE}^{(c)}$ and $\hat{\alpha}_{MAP}^{(c)}$. Let $\hat{\alpha}_{\bullet}$ be any of these estimators. We first show that a condition similar to (17) holds for $\hat{\alpha}_{\bullet}$. Since $\hat{\alpha}_{\bullet}$ and $\hat{\alpha}_{MPE}^{(J)}$ are strongly consistent,

$$\frac{1}{k} \left[L_k(\hat{\alpha}_{MPE}^{(J)}) - L_k(\hat{\alpha}_{\bullet}) \right] = \left(\log \hat{\alpha}_{MPE}^{(J)} - \log \hat{\alpha}_{\bullet} \right) \frac{1}{k} \sum_{i=1}^k \delta_{[n-k+i]} - \left(\hat{\alpha}_{MPE}^{(J)} - \hat{\alpha}_{\bullet} \right) \frac{1}{k} \sum_{i=1}^k \log e_{[n-k+i]}$$

converges a.s. to 0 as $n \rightarrow \infty$. Moreover, $\hat{\alpha}_{MPE}^{(J)}$ maximizes $L_k(\alpha)$, thus for k sufficiently large, we have:

$$\frac{1}{k} \left[L_k(\hat{\alpha}_{MPE}^{(J)}) - L_k(\hat{\alpha}_\bullet) \right] < \frac{\epsilon}{2},$$

where ϵ is as in (17). Now,

$$\frac{1}{k} [L_k(\alpha) - L_k(\hat{\alpha}_\bullet)] = \frac{1}{k} [L_k(\alpha) - L_k(\hat{\alpha}_{MPE}^{(J)})] + \frac{1}{k} [L_k(\hat{\alpha}_{MPE}^{(J)}) - L_k(\hat{\alpha}_\bullet)].$$

For k sufficiently large, $|\alpha - \hat{\alpha}_{MPE}^{(J)}| > \delta$ and $|\alpha - \hat{\alpha}_\bullet| > \delta$ hold simultaneously since $\hat{\alpha}_{MPE}^{(J)} - \hat{\alpha}_\bullet$ converges to 0 a.s.. Thus, for k sufficiently large, we have:

$$\sup_{|\alpha - \hat{\alpha}_\bullet| > \delta} \frac{1}{k} [L_k(\alpha) - L_k(\hat{\alpha}_\bullet)] < -\epsilon + \frac{\epsilon}{2} = -\frac{\epsilon}{2}.$$

Weak convergence of $\sqrt{k}(\alpha_0 - \hat{\alpha}_\bullet)$ can now be proved along the same lines as for $\hat{\alpha}_{MPE}^{(J)}$. Details are omitted. \square

Appendix B. Results for the GPD model.

Results are provided for a sample size $n = 1000$ and $m = 1000$ simulated samples. For each $\ell = 1, 2, 3, 4, 5$, we plot the empirical bias and RMSE of $\hat{\alpha}_\ell$ as a function of k for $p = 0.9, 0.75, 0.5, 0.25$ (Figure 7: $\alpha = 0.5$ and Figure 8: $\alpha = 0.75$).

In Table 6, we provide the averaged value (over the m simulated samples), empirical bias and RMSE of each $\hat{\alpha}_\ell$ at the optimal fraction level $k_{opt} = \arg \min_k RMSE[\hat{\alpha}_\ell(k)]$ (for $\alpha = 0.5$). Results for $\alpha = 0.75$ yield similar conclusions and are therefore omitted.

		k_{opt}	Estimator	RMSE	Bias
p=0.9	$\hat{\alpha}_1$	241	0.4871029	0.03316159	-0.012897068
	$\hat{\alpha}_2$	241	0.4848721	0.03397750	-0.015127864
	$\hat{\alpha}_3$	242	0.4937372	0.03125152	-0.006262834
	$\hat{\alpha}_4$	241	0.4871525	0.03301891	-0.012847470
	$\hat{\alpha}_5$	241	0.4849317	0.03383208	-0.015068322
p=0.75	$\hat{\alpha}_1$	197	0.4813512	0.04283156	-0.01864877
	$\hat{\alpha}_2$	174	0.4817153	0.04407021	-0.01828467
	$\hat{\alpha}_3$	229	0.4820538	0.03934719	-0.01794616
	$\hat{\alpha}_4$	175	0.4853335	0.04255187	-0.01466651
	$\hat{\alpha}_5$	174	0.4818332	0.04375507	-0.01816682
p=0.5	$\hat{\alpha}_1$	149	0.4650962	0.06667390	-0.03490379
	$\hat{\alpha}_2$	148	0.4593110	0.06983460	-0.04068903
	$\hat{\alpha}_3$	196	0.4655255	0.05698297	-0.03447449
	$\hat{\alpha}_4$	149	0.4654849	0.06588210	-0.03451512
	$\hat{\alpha}_5$	148	0.4597745	0.06900169	-0.04022548
p=0.25	$\hat{\alpha}_1$	114	0.4453977	0.09310355	-0.05460229
	$\hat{\alpha}_2$	150	0.4228396	0.10027361	-0.07716037
	$\hat{\alpha}_3$	144	0.4660381	0.07223843	-0.03396190
	$\hat{\alpha}_4$	114	0.4467343	0.09076946	-0.05326572
	$\hat{\alpha}_5$	150	0.4242566	0.09843514	-0.07574341

Table 6: Optimal results for GPD model with $n = 1000$, $m = 1000$ and $\alpha = 0.5$.

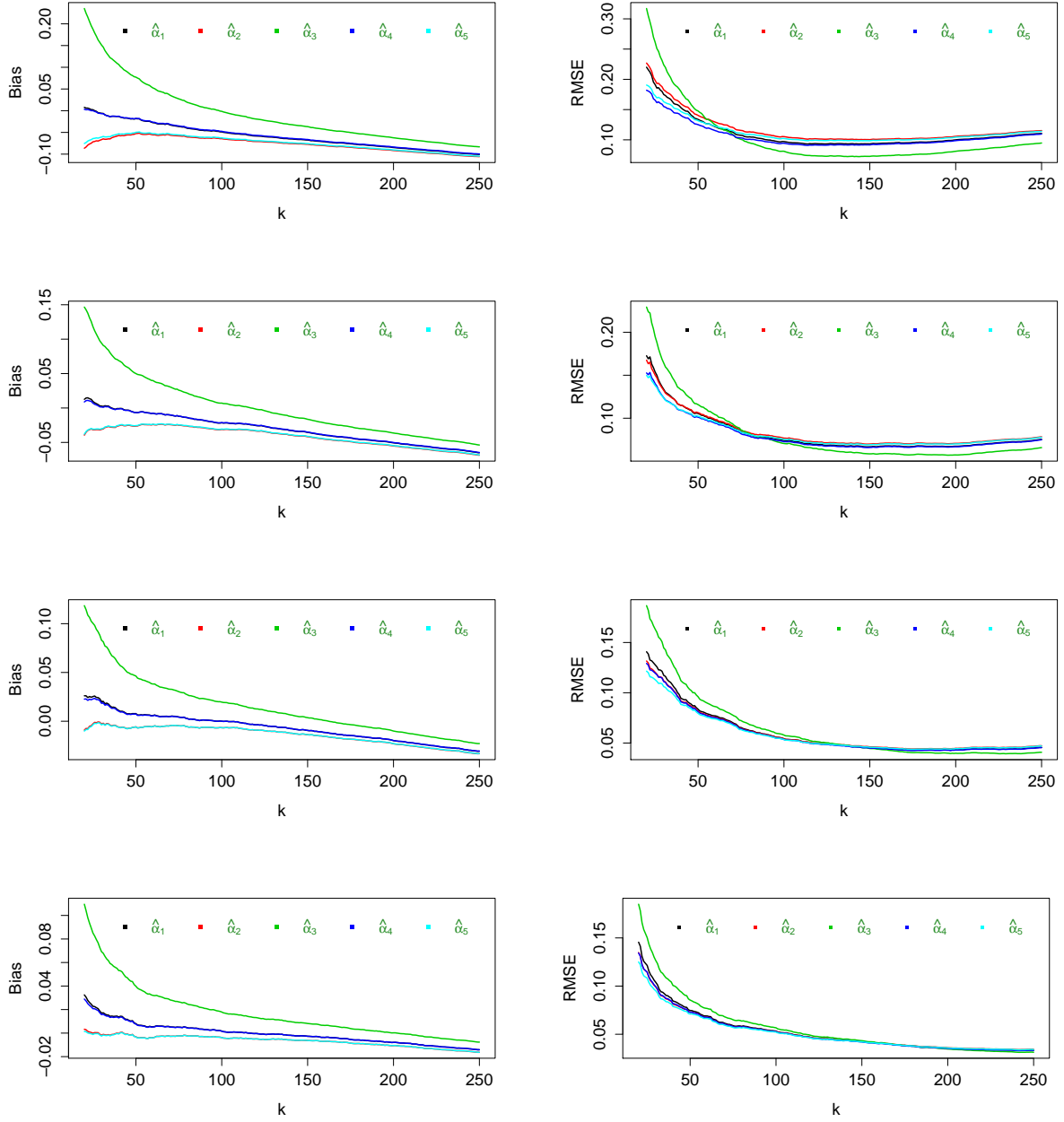


Figure 7: Empirical bias (left) and RMSE (right) of $\hat{\alpha}_l$ ($l = 1, \dots, 5$) for $\alpha = 0.5$ and $p = 0.25$ (1st line:), $p = 0.5$ (2nd line), $p = 0.75$ (3rd line) and $p = 0.9$ (4th line).

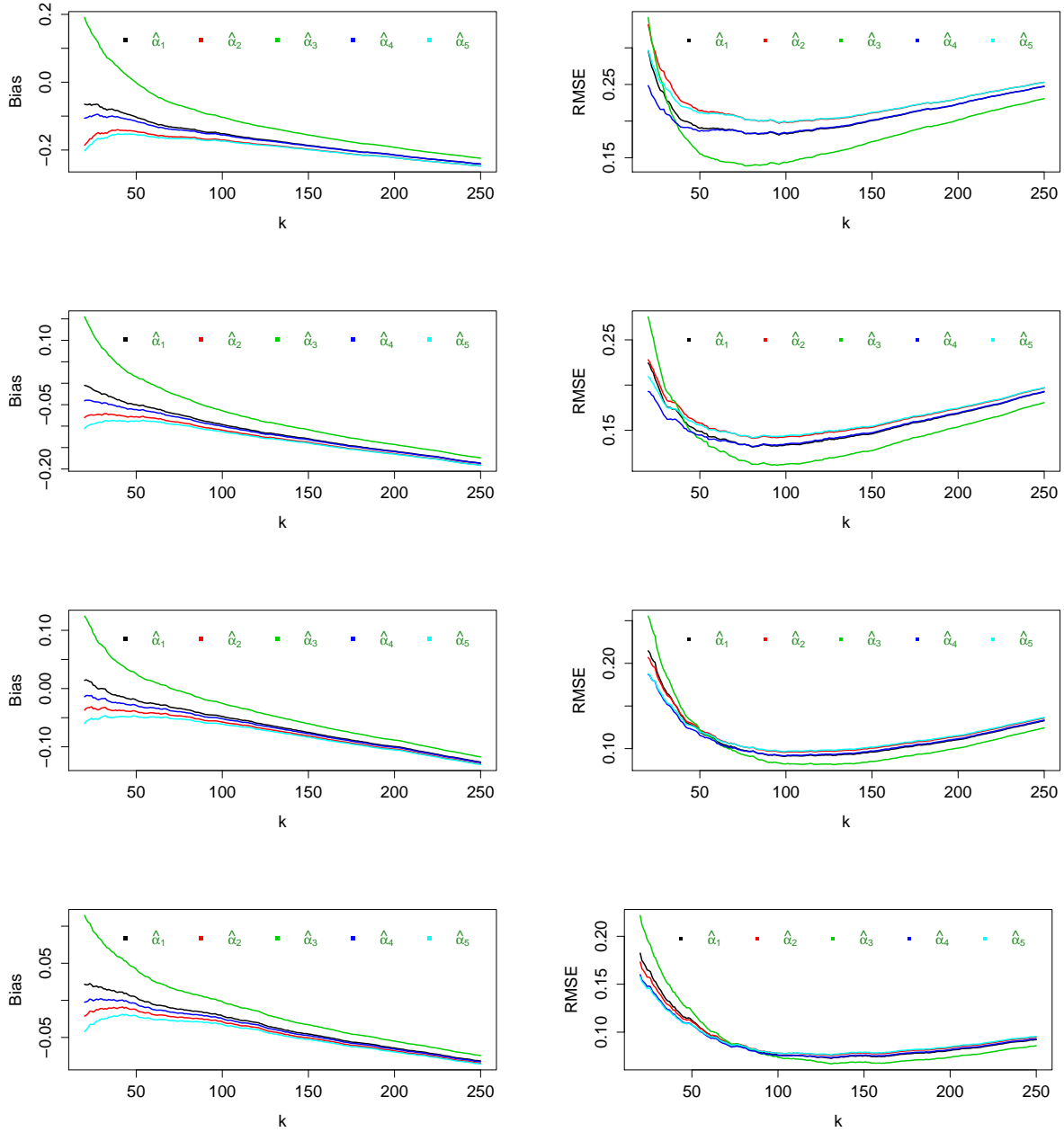


Figure 8: Empirical bias (left) and RMSE (right) of $\hat{\alpha}_l$ ($l = 1, \dots, 5$) for $\alpha = 0.75$ and $p = 0.25$ (1st line:), $p = 0.5$ (2nd line), $p = 0.75$ (3rd line) and $p = 0.9$ (4th line).

Acknowledgements

The authors acknowledge financial support from the CNEPRU project "Modélisation - Simulation Actuarielle de Risques (MSAR)" (project number B00220120012, funded by Agence Nationale de Développement de la Recherche Universitaire, Algeria) and from the project SESA (project number EDC25895, funded by Centre National de la Recherche Scientifique - CNRS, France - and Direction de la Post-Graduation et de la Recherche-Formation - DPGRF, Algeria). Part of this work was made while

A. Ameraoui was visiting INSA of Rennes and while J.-F. Dupuy was visiting the Université des Sciences et de la Technologie Houari Boumediene, Algiers.

References

- [1] Andersen P.K., Borgan O., Gill R.D., Keiding N., 1993. Statistical Models Based on Counting Processes. Springer-Verlag, New York
- [2] Beirlant J., Goegebeur Y., Segers J., Teugels J., De Waal D., Ferro C., 2004. Statistics of Extremes. John Wiley & Sons Ltd, Chichester.
- [3] Beirlant J., Guillou A., Dierckx G., Fils-Villetard A., 2007. Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes* 10(3), 151-174.
- [4] Brahimi B., Meraghni D., Necir A., 2013. On the asymptotic normality of Hill's estimator of the tail index under random censoring. Preprint: arXiv-1302.1666.
- [5] Cabras S., Castellanos M.E., 2011. A Bayesian approach for estimating extreme quantiles under a semiparametric mixture model. *Astin Bulletin* 41(1), 87-106.
- [6] Coles S., Powell E., 1996. Bayesian methods in extreme value modelling: a review and new developments. *International Statistical Review* 64(1), 119-136.
- [7] Deheuvels P., Haeusler E., Mason M., 1988. Almost sure convergence of the Hill estimator. *Mathematical Proceedings of the Cambridge Philosophical Society* 104(2), 371-381.
- [8] Dekkers A.L.M., Einmahl J.H.J., de Haan L., 1989. A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* 17(4), 1833-1855.
- [9] Dewan I., Prakasa Rao B.L.S., 1997. Remarks on the strong law of large numbers for a triangular array of associated random variables. *Metrika* 45(3), 225-234.
- [10] Diebolt J., El-Aroui M., Garrido M., Girard S., 2005. Quasi-conjugate bayes estimates for gpd parameters and application to heavy tails modelling. *Extremes* 8(1-2), 57-78.
- [11] Einmahl J.H.J., Fils-Villetard A., Guillou A., 2008. Statistics of extremes under random censoring. *Bernoulli* 14(1), 207-227.
- [12] Embrechts P., Kluppelberg C., Mikosch T., 1997. Modelling Extremal Events. Springer-Verlag, Berlin.
- [13] Ergashev B., Mittnik S., Sekeris E., 2013. A Bayesian approach to extreme value estimation in operational risk modeling. *Journal of Operational Risk* 8(4), 55-81.
- [14] Ghosh J.K., Delampady M., Samanta T., 2006. An Introduction to Bayesian Analysis. Springer, New York.
- [15] Gomes M.I., Neves M.M., 2011. Estimation of the extreme value index for randomly censored data. *Biometrical Letters* 48(1), 1-22.
- [16] Gomes M.I., Oliveira O., 2003. Censoring estimators of a positive tail index. *Statistics & Probability Letters* 65(3), 147-159.
- [17] de Haan L., Ferreira A., 2006. Extreme Value Theory. Springer, New York.
- [18] Hill B.M., 1975. A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 3(5), 1163-1174.
- [19] Jeffreys H., 1961. Theory of Probability. Oxford University Press.

- [20] Liang Z.M., Li B.Q., Yu Z.B., Chang W.J., 2011. Application of Bayesian approach to hydrological frequency analysis. *Science China Technological Sciences* 54(5), 1183-1192.
- [21] Mason D.M., 1982. Laws of large numbers for sums of extreme values. *Annals of Probability* 10(3), 754-764.
- [22] do Nascimento F.F., Gamerman D., Lopes H.F., 2012. A semiparametric Bayesian approach to extreme value estimation. *Statistics and Computing* 22, 661-675.
- [23] Pickands J., 1975. Statistical Inference Using Extreme Order Statistics. *Annals of Statistics* 3(1), 119-131.
- [24] R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [25] Resnick S.I., 2007. Heavy-Tail Phenomena. Probabilistic and Statistical Modeling. Springer, New York.
- [26] So M.K.P., Chan R.K.S., 2014. Bayesian analysis of tail asymmetry based on a threshold extreme value model. *Computational Statistics and Data Analysis* 71, 568-587.
- [27] Stephenson A., Tawn J., 2004. Bayesian inference for extremes: accounting for the three extremal types. *Extremes* 7(4), 291-307.
- [28] van der Vaart A.W., 1998. Asymptotic Statistics. Cambridge University Press, Cambridge.
- [29] Venables W.N., Ripley B.D., 2002. Modern Applied Statistics with S. Springer, New York.
- [30] de Zea Bermudez P., Kotz S., Parameter estimation of the generalized Pareto distribution-Part II. *Journal of Statistical Planning and Inference*, 2010. **140(6)**: p. 1374-1388.
- [31] Zellner A., 1971. An Introduction to Bayesian Inference in Econometrics. John Wiley & Sons Ltd, New York.
- [32] Zellner A., 1996. Models, prior information, and Bayesian analysis. *Journal of Econometrics* 75(1), 51-68.